

Political History in the Digital Age: The challenges of archiving and analysing born digital sources.

The vast bulk of source material for historical research is still paper-based. But this is bound to change. Dr Helen McCarthy considers the lessons from the Mile End Institute's conference on Contemporary Political History in the Digital Age. The specific challenges of using a 'born digital source' is an area that requires considerable attention. For political historians, the advent of 'e-government' and personal digital archives, and the many formats and artefacts involved, is thrilling but also intimidating.



Historians like digging around in archives.

The materiality of the primary source is part of the allure of historical research: rummaging through dust-covered files, turning the decomposing pages of thick-bound volumes, removing rusty paperclips, perusing bundles tied with ancient string – it's all part of the voyage of discovery into the past which drew most of us to our careers as historians.

The digitisation of many paper-based sources in recent years has altered that landscape considerably. Now, depending on subject area, 'archival' research just as often involves staring at computer screens and clicking on a mouse as it does handling physical records. This shifting reality is reflected in the kinds of training that postgraduate research students habitually receive: the Institute of Historical Research, for instance, runs a popular course on ['Internet for Historical Research'](#) in which participants learn about advanced digital search techniques and electronic referencing tools, as well as how to navigate the wealth of primary and secondary source material available online.

But for many historians, the more specific challenge of using the 'born digital source' – that is, a source which only exists in digital form – is not one to which they have given much thought. This is despite a huge amount of valuable activity and conversation taking place amongst professional archivists and information management specialists in government.

Disruptive challenge

The Mile End Institute in partnership with the Foreign Office Historians and sponsored by the British Academy hosted an event in February 2016 on [Contemporary Political History in the Digital Age](#) in order to address this issue and bring these groups into a fruitful dialogue about the future of the digital archive.

"Leafing through an individual's collected correspondence or handwritten diary is one thing; making sense of endless email tails, analysing documents in multiple electronic iterations...is quite another."

- Dr Helen McCarthy

Part of the reason why historians have been late to the party is because for all but the most contemporary of scholars, the vast bulk of relevant source material is still paper-based. But as researchers start to think historically about the 1990s and 2000s, that will inevitably change. To take the example of political history: since the advent of 'e-government' in the late 1990s, the business of the central and local state has increasingly taken digital form, both in terms of internal communications and delivering public services to citizens.

In the future, the buff-coloured files containing neatly organised papers that we are used to reading at The National Archives will simply disappear. [Russell Davies](#), former head of the Government Digital Service, painted an exciting picture of the burgeoning digital state, in which civil servants share ideas by using web tools like [Slack](#), [Trello](#), [Basecamp](#) and [Pivotal Tracker](#), whilst ordinary citizens co-create services by building improved code for government websites using sites such as [GitHub](#). Davies highlighted the disruptive challenge which this kind of distributed activity poses to the traditional policy-making model, in which ministers set priorities, senior civil servants come up with options, and lower-ranking officials get on with the more prosaic task of 'delivery'.

But what might this mean for future historians who want to reconstruct the policy decision-making process? Will we have access to those power-point slides, digital conversations and citizen-generated innovations which Davies sees as having a transformative effect on the very nature of the state? Do we in fact need to write a comprehensive history of technology in government since the 1980s, documenting exactly how different technologies have been adopted and deployed in different parts of the state, and what impact they have made on administrative processes and institutional cultures?

Twitter pitfalls

Another major challenge concerns the future reconstruction of the public sphere of political debate and activism. [Katrin Weller](#), an information scientist who has worked on the Library of Congress's Twitter archive, described not only the considerable technical difficulties of preserving user-generated, social media content, but the intellectual and methodological pitfalls facing any researcher who wishes to work with such sources.

To analyse a tweet meaningfully, we must understand its context in depth: who tweeted it and when did it appear? What webpages did it link to? Was it tweeted from a mobile device, a laptop or a desktop, and does that make any difference to how we interpret its meaning? Who was the tweet aimed at? Who read it? Who replied, retweeted or favorited it? Did it include a hashtag, and if so, do we understand its significance? Can we look at all other tweets

with the same hashtag? Can we capture retrospectively the dynamic interactivity inherent in a medium such as Twitter?



Image credit: Sediment Core Repository by Hannes Grobe, AWI (curator of archive)

James Baker, a digital historian based at the University of Sussex (whose speaking notes can be viewed [here](#)), complicated things further by reflecting on the massive volume of unpublished born digital sources which are stored on old computer hard drives and memory sticks belonging to private individuals. As Baker pointed out, such artefacts can contain hundreds of thousands of files comprising terabytes of data, and include complex components – from semi-automated browser caches to file use metadata and downloads folders.

Digital forensics

The prospect of working with these ‘personal digital archives’ is thrilling but also intimidating.

Leafing through an individual’s collected correspondence or handwritten diary is one thing; making sense of endless email tails, analysing documents in multiple electronic iterations, and reconstructing viewed webpages in their original (but quite possibly now defunct) browsers, is quite another. Will historians have to become experts in ‘digital forensics’, adopting methods currently pioneered by law enforcement authorities, to access and interpret this kind of data? And by what legal mechanisms do these ‘digital assets’ pass into the public domain after an individual’s death, thus becoming available for researchers to plunder?

As well as raising important technical issues and ethical questions, what Weller and Baker’s contributions both point to is the importance of historians acquiring a sophisticated awareness of the place of digital technologies in people’s day-to-day lives. Just as we need to understand how government has used technology over time, we must understand change and continuity in everyday social practices – from browsing the internet and communicating with social media, to taking photos or ordering groceries on a smartphone.

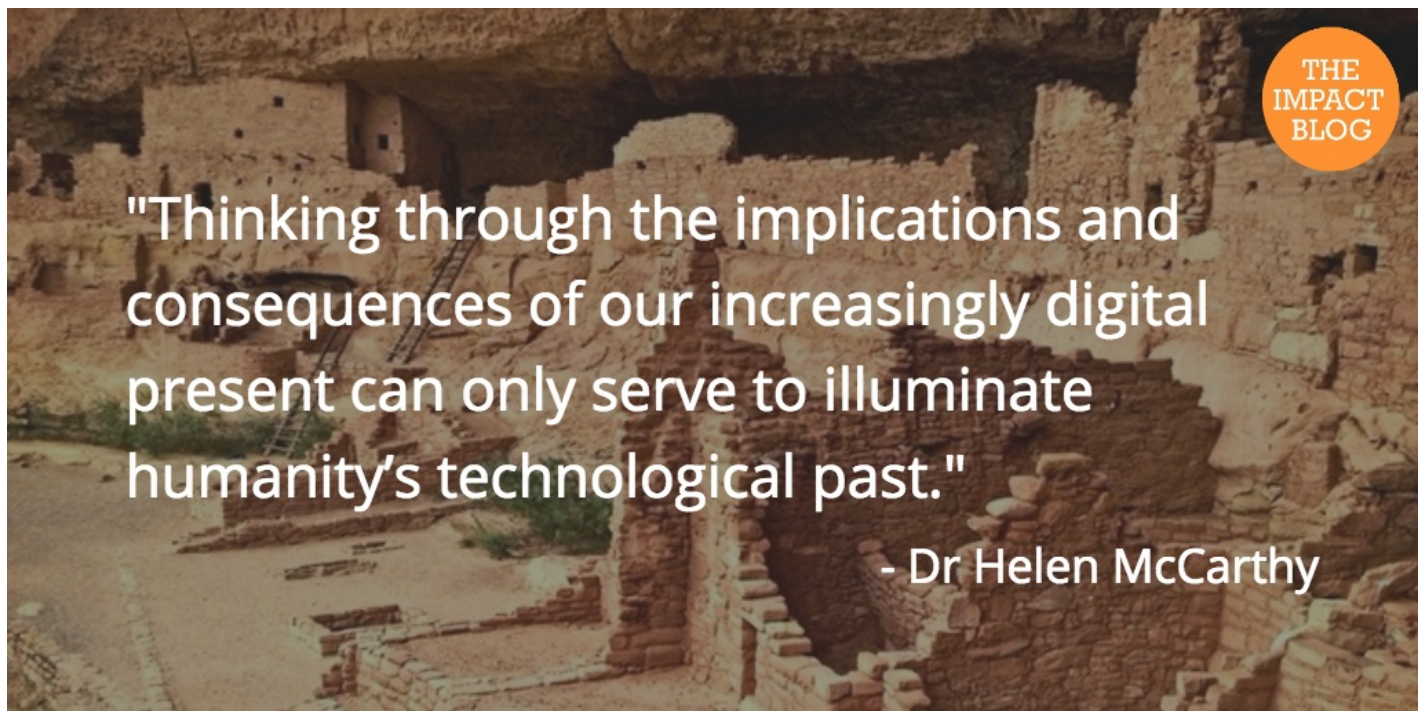
Without reproducing hyperbolic narratives of ‘revolutionary’ change, we must nonetheless be sensitive to the distinctive characteristics of some digital sources: a blog post by a soldier serving in Afghanistan containing multiple links and generating reader comments is not, as Katrin Weller noted, the same as a private letter sent home from the trenches of the First World War.

Similarly, an email or text message is often closer to the spoken rather than written word. Here, as well as drawing directly on the raw metadata produced by our digital lives (what times of day do we use twitter? How long do we surf the internet for? When do we switch off our phones?), historians can deploy familiar oral history methods but also learn much from interdisciplinary work going on in this field, such as that carried out at the [Centre for Creative and Social Technologies](#) at Goldsmiths College, the [Centre for Social Media Research](#) at the University of Westminster, and the [Centre for the Analysis of Social Media](#) at the think tank Demos.

Professional sources

Professional archivists have, of course, been engaged in thinking about born digital sources for some time. [Simon Demissie](#) offered an overview of how The National Archives is grasping the challenge, from working with government to develop a digital records strategy (and, importantly, appraisal policy) to creating the [UK Government Web Archive](#), which comprises historical government websites, twitter accounts and YouTube content. This adds to the [UK web archive](#) which the British Library has been creating since 2013, when its legal deposit remit was extended to include the world wide web, a large subset of which is freely available online. And it complements the numerous digital archiving initiatives which have emerged in recent years on a voluntary and non-profit basis, such as the 150 billion historical web pages which form part of the [Internet Archive](#).

Will these platforms capture everything for posterity? Inevitably no. Historians always wring their hands over gaps in the archive – the letters that were burned, the files that were thrown away – but, in truth, historical research always involves coming to terms with loss.



Perhaps the digital age fosters an illusion of the possibility of permanence, but the fact is that our data in the future will be partial just as it has been in the past. As a result, historians will need to transpose their long-established disciplinary skills and instincts into a digital register: asking the usual critical questions about their source material – how it was produced and why it has survived – and establishing a deep and rich set of contexts through which to interpret it.

For me, the ultimate lesson from the event is that historians just have to – well, just go out and start doing it. To grasp the full possibilities and challenges of the digital archive, we need to begin to work with born digital sources. Some are already doing so: the FCO's Chief Historian, Patrick Salmon, presented the results of an experiment in which he and members of his team compiled a set of primary source documents pertaining to the G8

Summit in 2005, using the digital file structure which was introduced to replace the old paper registration system. The process revealed the crucial importance of search engine software, which must be intelligent enough to reassure researchers that they have, indeed, found all relevant material. As Salmon remarked, under the old system, one knew one had seen all the documents when one reached the end of the physical file, and the file's location within a hierarchy of records was crystal clear.

New opportunities

So should historians learn to code in order to develop their own search tools? At the very least it seems that they should have a much stronger understanding of how search engines work and be able to critique their functions from an informed standpoint.

There are many other exciting initiatives on-going in the UK: the [Big UK Domain Data for the Arts and Humanities](#), for instance, is a project funded by the AHRC which promotes innovative methodologies, including historical ones, for studying web data created between 1996 and 2013. The AHRC is also funding a new network, [Born Digital: Big Data and Approaches for History and the Humanities](#) which will specifically explore the problems of born digital data from a historical perspective.

Undergraduate and postgraduate programmes in history are beginning to introduce students to the realities of research in a digital world, whilst working in interdisciplinary teams across the humanities and computer sciences, although not a wholly new concept for historians, has become well-established in the Digital Humanities departments springing up across UK universities.

If the future is digital, then it is also bright. Historians have a lot to gain from engaging with the challenge of the born digital source, including scholars whose interests do not lie in the late 20th or early 21st centuries. Understanding technology as a driver of change in human societies is a task for historians of all periods and all civilisations. Thinking through the implications and consequences of our increasingly digital present can only serve to illuminate humanity's technological past.

This piece was first published on the [Mile End Institute blog](#) and is reposted with the author's permission.

Note: This article gives the views of the author, and not the position of the LSE Impact blog, nor of the London School of Economics. Please review our [Comments Policy](#) if you have any concerns on posting a comment below.

About the Author

Helen McCarthy is Senior Lecturer in History at Queen Mary University of London and Deputy Director of the Mile End Institute. She organised the conference, [Contemporary Political History in the Digital Age](#), in partnership with the Foreign and Commonwealth Office Historians, as part of a wider programme of activity on the theme of Rethinking Contemporary British Political History, which is funded by a British Academy's Rising Star Engagement Award

- Copyright 2015 LSE Impact of Social Sciences - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.